RANDOM PROCESSES IN PHYSICS AND COMMUNICATIONS

SOME TOPICS IN INFORMATION THEORY

C. E. Shannon

Previous work in communication theory [2] has shown that *amount* of information for purposes of communication has a natural measure in terms of entropy type formulas $H = -\Sigma p \log p$. This has led to theorems giving the most efficient encoding of the messages produced by a stochastic process into a standard form, say a random sequence of binary digits, and for the most efficient use of an available communication channel. However, no concept of information itself was defined. It is possible to formulate an approach to the theory in which the information sources in a communication network appear as elements of a lattice.

The leading idea is that any reversible translation of the messages produced by a stochastic process, say by a non-singular finite state transducer, should be regarded as containing the same information as the original messages. From the communications point of view, knowledge of the Morse code translation of the text originating at a telegraph office is equivalent to knowledge of the text itself. Thus we consider the information of a source to be the equivalence class of all reversible translations of the messages produced by the source. Each particular translation is a representative of the class, analogous to describing a tensor by giving its components in a particular coordinate system.

Various theories may be obtained depending on the set of translation operations allowed for equivalence. Two choices lead to interesting and applicable developments: (1) the group of all finite state transducers (allowing effectively positive or negative delays), (2) the group of *delay free* finite state transducers, in which it is required that the present output symbol be a function of the present and past history of the input, and similarly for the reverse transducer.

The first case is the simplest and relates most closely to previous work in which unlimited encoding delays at transmitter and receiver were allowed. A transitive inclusion relation between information elements, $x \geqq y$, (inducing a partial ordering) means that y can be obtained by operating on x with some finite state transducer (not necessarily reversible). The entropy of a source (which is invariant under the group of reversible transducers) appears as a norm monotone with the ordering. The least upper bound for two elements is the total information in both sources, a representation being the sequence of ordered pairs of letters from the two sources. A greatest lower bound can also be defined, thus resulting in an information lattice. There will always be a universal lower bound, and if the set of sources considered is finite, a universal upper bound. The lattices obtained in this way are, in general, non-modular. In fact, an information lattice can be constructed isomorphic to any finite partition lattice.

A metric can be defined by $\rho(x, y) = H_x(y) + H_y(x)$ satisfying the usual

requirements. This introduces a topology and the notion of Cauchy convergent sequences of information elements and of limit points. If convergent sequences are annexed to the lattice as new points, with corresponding modifications of the definition of equality, etc., there result continuous lattices, for example the set of all the abstractions of the total information in the system by finite state transducers, or limiting sequences of such transducers.

The delay free theory leads also to a lattice but the problems, while perhaps more important in the applications, are less well understood. The entropy of a source is no longer sufficient to characterize the source for purposes of encoding, and in fact an infinite number of independent invariants have been found. Certain of them are related to the problem of best prediction of the next symbol to be produced, knowing the entire past history. The delay free theory has an application to the problem of communication over a channel where there is a second channel available for sending information in the reverse direction. The second channel can, in certain cases, be used to improve forward transmission. Upper bounds have been found for the forward capacity in such a case. The delay free theory also has an application to the problem of linear least square smoothing and prediction [1]. A minimum phase filter has an inverse (without delay) and therefore belongs to the delay free group of translations for continuous time series. The least square prediction problem can be solved by translating the time series in question to a canonical form and finding the best prediction operator for this form.

References

1. H. W. Bode and C. E. Shannon, *A simplified derivation of linear least square smoothing and prediction theory*, Proceedings of the Institute of Radio Engineers vol. 38 (1950) pp. 417–425.

2. C. E. Shannon and W. Weaver, *A mathematical theory of communication*, University of Illinois Press, 1949.

Bell Telephone Laboratories, Inc.,
Murray Hill, N. J., U. S. A.